DISCUSSION FOR DESIGNING CLINICAL PROGRAMS FOR DEVELOPING DRUGS, DEVICES, OR BIOLOGICAL PRODUCTS INTENDED FOR THE TREATMENT OF RHEUMATOID ARTHRITIS (RA)

TABLE OF CONTENTS

DISCUSSION FOR DESIGNING CLINICAL[1] PROGRAMS FOR DEVELOPING DRUGS, DEVICES, OR BIOLOGICAL PRODUCTS INTENDED FOR THE TREATMENT OF RHEUMATOID ARTHRITIS (RA)

"This document is intended as a basis for discussion at the 7/23/96 workshop on RA product development. It is intended as a starting point, not as a definitive document."

I. POTENTIAL CLAIMS FOR THE TREATMENT OF RA

Although label claims have diverse legal and regulatory ramifications, their central purpose is to inform prescribers and patients about the documented benefits of the product. Because RA is a chronic, symptomatic disease that can result in a variety of adverse outcomes with different chronology, severity, and overall patient impact, it has been difficult to develop a common view of outcomes that could be the basis for distinct claims. The following list of possible claims represents the current views of Agency rheumatologists about achievable and clinically relevant overall outcomes.

---

[1] (When Finished) -- This guidance has been prepared by the Rheumatology Working Group of the Medical Policy Coordinating Committee (MPCC) of the Center for Drug Evaluation and Research (CDER) in collaboration with the Center for Biologics Evaluation and Research (CBER) and the Center for Devices and Radiological Health. Although this guidance does not create or confer any rights for or on any person and does not operate to bind FDA or the industry, it does represent the agency's current thinking on the evaluation of drugs, devices and biological products intended for the treatment of Rheumatoid Arthritis. For additional copies of this guidance contact the Division of Communications Management (formerly the Executive Secretariat Staff), HFD-210, Center for Drug Evaluation and Research, FDA, 5600 Fishers Lane, Rockville, MD 20857 (Phone: 301-594-1012). An electronic version of this guidance is also available via Internet using FTP, Gopher or the World Wide Web (WWW). For FTP, connect to the CDER anonymous FTP server at CDVS2.CDER.FDA.GOV and change to the "guidance" directory. For Gopher connect to the CDER Gopher server at GOPHER.CDER.FDA.GOV and select the "Industry Guidance" menu option. For WWW, connect to the FDA Home Page at WWW.FDA.GOV and go to the CDER section.

These claims would be based on what was demonstrated in the trials, and would not segregate agents into particular classes: rather, claims could be accumulated by a particular agent as they were demonstrated.

A. Reduction in the signs and symptoms of RA

This claim is intended to reflect symptomatic benefit, or benefit that includes improvement in signs of disease activity as well as symptoms. Ordinarily such a claim would be established by trials with exposure no shorter than 12 weeks (**Question for discussion: is 6 months more appropriate**?) in duration, and the claim is intended to contain a statement about the period of symptomatic control demonstrated. Sample outcome measures and trial designs that would support claim A include:

1. Composite endpoints or "indices" in placebo or active-controlled trials. These composites can be used in trials to define by-patient success or failure, i.e., use of the ACR preliminary definition of improvement.[1] [2] Timing of the measurements and how to count dropouts are examples of issues that should be resolved in designing these trials.

2. Well accepted sets of sign/symptom measures, eg, ACR core set, used as outcome measures. The criteria for "success" and the methods for statistical analysis should be prospectively defined and agreed upon by Agency and sponsor.

3. FDA's traditional measures of joint counts (pain and tenderness) and global assessments (physician and patient) in placebo or active controlled trials. Typically, a statistically significant difference compared to placebo in 3 of the 4 measures has been used as the criterion for a successful placebo controlled trial; for comparative trials equivalence on 3 of the 4 measures would be used. The trial design section contains comments on the issues of repeated measures, handling of dropouts, equivalence trial analysis, and suggested analytical techniques for trials using these measures.

Historically, shorter term trials in RA have used repeated measures, often at biweekly or monthly intervals, to assess patient status. Since the evaluation of short term benefit is primarily focused on alleviation of current symptoms, it seems logical to use

a statistical method that incorporates the treatment effect at every evaluation, and not put exceptional weight on the last evaluation, for A2 and A3. This approach could help minimize the impact of dropouts.

B. Remission[3]

The claim of "remitting agent" is intended to reflect substantial therapeutic activity, of greater import than the relief of signs and symptoms of RA. It is proposed that a remission claim be based on a statistically significant improvement in remission rate, ("remission" defined by 1981 ACR criteria), sustained over 6 months, that is at least 20% higher in absolute value than the observed rate in control patients treated with current standard therapies appropriate for their stage and severity of disease. ***(Issues for panelists: in what circumstances should this number be different?)*** *It is anticipated that trials intending to evaluate remission be at least 1 year in* duration.

C. Prevention of structural damage

Prevention of structural damage is clearly an important goal of RA therapy. It is anticipated that trials evaluating this claim be at least one year in duration.

The OMERACT composite or similar measure evaluating primarily current symptomatic status should also be evaluated in trials evaluating structural progression. Lack of success in current disease control in a trial that demonstrated, for example, slowing of radiographic progression, would raise concerns about the type of claim granted the agent, and the overall plausibility of the finding, that should be addressed.(Issue for panelists: ***What concomitant clinical claims are needed?***)

The following are examples of outcome measures that could be used to support prevention of structural damage claims. Because of the paucity of agents known to be effective in preventing progression, trials should be designed to show superiority rather than equivalence.

1. Retardation of X-ray progression - using either the Larsen, modified Sharp, or other index. Radiographic claims should be based on comparisons of films taken one year (or longer) after baseline to baseline films in all randomized patients with a

retention rate of at least 85% to reduce the bias from dropouts. (***Question for panel - Is 85% a problem?)***

2. Prevention of new X-ray erosions - maintaining an erosion-free state or preventing new erosions. This claim would have the same type of data set outlined for #1.

3. Other sensitive measures, e.g., MRI, could be evaluated. Because of the potentially great sensitivity to change, the magnitude of difference generally considered to be clinically significant (rather than simply statistically significant) should be determined in advance as part of the trial design.

D. Prevention of disability An appropriate design here might be a study using time to progression of disability. ***(Panel - What concurrent traditional endpoint improvement is necessary, if any? What measures of disability and how much of a difference should be shown?)***

E. Improvement (or prevention of deterioration) in quality of life

Improvement in quality of life is envisioned as an adjunctive claim that would be predicated on also substantiating one of the above claims(A, B, or C.) Either general health assessments or more specific disease related quality of life instruments would be acceptable. Currently, designs intended to show superiority would be used to support quality of life claims because there are no products approved with this claim. It is anticipated that an area under the curve or trend analytic approach would be used for statistical analysis. ***(Panel - comments on duration of trials or effect?)***

II. CONSIDERATIONS IN RA PRODUCT DEVELOPMENT

Most of the following information on preclinical and early clinical development pertains primarily to pharmaceuticals (drugs and biologicals.) For information on early development of medical devices, please refer to (ref) and to the later section on devices in this guidance. The information on efficacy evaluation is relevant to most treatments.

Certain problems recur in RA product development. Frequently encountered issues include: (1) selecting of appropriate in vitro and animal model activity screening; (2)

"orchestrating" the preclinical/clinical interface; (3) balancing the potential need for intervention early in the disease course with the need to avoid exposing patients with mild disease to toxic agents; (4) understanding drug co-administrations and combination therapies; (5) designing adequate and practical long-term safety monitoring; and (6) definitively showing efficacy. The following sections discuss approaches to these challenges.

A. Pharmacology Strategies

This section will focus on preclinical issues that are specific to the clinical development of anti-rheumatic therapeutics. In designing toxicity studies, and the timing of such studies, industry should consult the current recommendations and guidelines that exist for drugs, devices and biological products. In addition, ICH guidelines currently under development, e.g., "Timing of preclinical studies to support clinical trials" or "Principles for the development of biotechnology-derived products," will be helpful once they are available.

1. Preclinical pharmacokinetics

Information on the absorption, distribution, metabolism and excretion are necessary during the IND phase but need not all be completed prior to phase 1. Generally, for initial studies in humans, determination of pharmacokinetic parameters such as AUC, Cmax, and t1/2 will be sufficient.

In some cases, especially where there is cause for concern regarding potential effects (i.e., interactions) with concomitant therapies, information on impact of such therapies on PK parameters may be needed to optimize dosing regimens and/or assessing safety. For example, drug/drug and drug/biologic interactions may be examined in an in vitro system using human hepatocytes. Similarly, competitive binding and displacement of a coadministered drug at the plasma protein site may be examined in in vitro studies using human plasma proteins.

2. Preclinical biological activity (Pharmacodynamics)

The biological activity or potency of a potential rheumatic therapy or therapies should be established using one or more in vitro, in vivo or ex vivo preclinical model systems.

3. In vitro

Data from in vitro studies can be useful not only in defining the potential mechanism of a drug or biological but also to determine relevance of a particular animal species proposed for in vivo assessment of activity or safety. These data are especially useful if a potential surrogate marker can be identified and subsequently used as a surrogate marker in preclinical and/or clinical studies to assess activity or safety.

Several in vitro tests may be developed depending on the mechanism of action of the drug or biologic. For example, binding assays may be useful for developing antagonists to class II MHC and cytokines. In vitro functional assays, e.g. platelet and neutrophil aggregation, are useful tests for developing inhibitors of inflammatory mediators. Sponsors are encouraged to develop in vitro functional assays for the antigen presentation and recognition cascade that would help in developing antagonists to TCR for CD4+ or CD8+ T-cells or MHC antagonists for the antigen presenting cells.

4. In vivo

Relevant species and/or animal model(s), that mimic one or more aspects of rheumatic disease in target population, should be used. Selection of animal models of disease should be considered not only in the analysis of potential pharmacodynamic responses but also to assess mechanism-based toxicity. Several factors are considered when selecting animal models. Ideally, products that are targeted for a specific subset of the population should be developed in an experimental model(s) that is most responsive to the marker for the subset of the population. For example, the rat model is not sensitive to inhibitors of 5-lipoxygenase products for an in vivo pharmacodynamic response. Therefore, the mouse or rabbit model should be used to evaluate the anti-inflammatory activity of leukotriene inhibitors.

Some experimental animal models that are used to screen for potential therapeutic candidates are discussed below: for a more detailed discussion please refer to the review by Mukherjee et al.[4]

Naturally occurring arthritis or autoimmune response in experimental animals (e.g., MRL/1 mice, Biozzi H mice, DBA/1 mice etc.)

These models are useful for examining the onset of drug-induced tolerance and the effect of the drug on an organ system involved in the autoimmune processes. The MRL/1 mouse model is useful to predict the effect(s) of a drug in lupus. Sponsors are encouraged to develop similar disease models that would allow screening of drugs on selective organs or sites involved in rheumatoid arthritis.

The MRL/1 mouse is a useful model for evaluating immunosuppressants and hormones in lupus erythematosus.

Rat carrageenin-induced acute model of inflammation

This model is useful to assess the potential activity of cyclooxygenase inhibitors. There are several reports suggesting that a good correlation exists between the activity of NSAIDs in this model and their clinical efficacy. The usefulness of the model for developing isozyme selective COX-II inhibitors should be analyzed in the future.

Adjuvant induced arthritis in rats:

Adjuvant induced arthritis in rats has been used extensively for screening antiarthritic drugs since the early sixties. However, other than NSAIDs, clinically effective anti-rheumatic drugs, i.e., gold salts, penicillamine and sulfasalazine have not shown impressive activity in this model. Immunomodulatory drugs have shown activity in this model. Therefore, drugs that potentially involve regulation of cytokine expression, as well as antimetabolite-like immunosuppressants, are candidates for evaluation in rat AA.

Streptococcal cell wall induced arthritis:

IL-1 receptor antagonist showed activity in this model, thus it may be used for developing cytokine inhibitors. The model may also provide further insights into the mode of action of immunomodulatory drugs.

Collagen-induced arthritis (CIA)

Collagen-induced arthritis is often considered a valid model for human RA because of the involvement of localized major histocompatibility complex class II-restricted T-helper-cell activation, and similar histopathological lesions. Radiographs of joints affected by CIA often show erosive changes similar to those seen in human RA and progressive arthritis often results in RA-like joint deformity and dysfunction. As occurs in some patients with human RA, anti-collagen antibodies develop in the CIA model.

The collagen-induced arthritis model may be useful for developing immunosuppressants and steroid hormones as well as inhibitors of inflammatory mediators, e.g. cytokines, prostaglandins, etc. The routinely used antirheumatic drugs e.g., gold, penicillamine, chloroquine have questionable activity in the collagen induced arthritis model. However, the CIA has merit over the adjuvant model with respect to the involvement of synovial tissues, and it can be induced in several animal species. Therefore, this model may be useful for evaluating drugs that respond species-selectively, e.g., leukotriene antagonists and 5-lipoxygenase inhibitors. This model should be explored further for evaluation of antiarthritic drugs that contribute to the remission of the disease. It is also suggested that incorporation of a functional parameter e.g., ability of walking or climbing, etc., would facilitate the screening of long acting antirheumatic drugs. Currently, these functional tests are not routinely used for preclinical testing. Therefore, development of additional parameters in arthritic animals is encouraged.

Experimental organ transplant in animals:

Effect of the drug in this model can provide evidence of activity, as well as safety for a target organ. This model provides a reliable short-term in vivo test for evaluating immunosuppressants and antimetabolites.

Model to assess the anti-phlogistic response:

Most of the animal models that involve inflammation in the paw may be used for measuring antiphlogistic action of a drug.

5. Transgenic animal models

   A number of transgenic animal models for the study of rheumatoid arthritis are being developed and may prove useful over the next decade. Some examples include:

   Transgenic mice that carry the env-pX region of the human T cell leukemia virus type I genome.

   TNF-transgenic mouse model.

   Transgenic mice that over-express or delete certain genes continue to be developed as in vivo models to study numerous aspects of the pathophysiology of RA such as those involving cytokines, enzymes and cell adhesion molecules. It is anticipated that these transgenic models will be increasingly useful preclinical models in the future.

6. Ex vivo

   Evaluation of inhibition of cyclooxygenase and phospholipase catalyzed products in mononuclear and platelet cells have been used to study the selectivity of a product to cyclooxygenase isozymes.

7. Preclinical toxicology

   Characterization of the general toxicity of a drug or biological product should be performed in accordance with the proposed clinical use.

   The intended use of the product determines whether specialized toxicity studies such as reproductive toxicity, genotoxicity and carcinogenicity are needed. It should be noted that the prevalence of autoimmune disease is especially high in females. Fertility and embryo toxicity studies should be completed early in clinical development to support the inclusion of women in early phases of clinical trials.

   Of particular concern are the immunomodulatory or immunosuppressive drugs and biologics intended to be

administered to arthritic patients either as monotherapy or in combination with known antiarthritic drugs, e.g., methotrexate. These biological activities raise additional questions about the possible delayed effects and reversibility of the immunosuppression. For such agents it is important to establish a dose response for the immunotoxic potential of the single agent in an experimental model of disease, if available, or in a relevant animal species.

In those cases where the margin of safety is small, and/or there are profound changes in immune function parameters or clinical markers of immunosuppression, e.g., lymphocyte counts, CD4 counts, TNF levels or MLR tests, it is recommended that potential drug-drug or drug-biological interaction be investigated preclinically to establish whether there is an additive or synergistic effect with products likely or intended to be co-administered with the investigative product. In such cases it will be important to use an in vivo model that is relevant and sensitive to both agents being evaluated. Pharmacokinetic studies may also be useful to help predict whether there may be enhanced toxicity due to concomitant or combination therapies.

B. PHARMACOKINETIC / PHARMACODYNAMIC STRATEGIES

1. PK/PD STUDIES - The types of pharmacokinetic studies for the development of any anti-rheumatic compound can be broken down into two broad categories: *in vivo* and in vitro studies required for all drugs, and *in vivo* studies specific to the anti-rheumatic category.

    a. *in vivo* pharmacokinetic studies for new molecular entities:

        1) Radiolabel studies or other suitable methodology - To establish the disposition and metabolic fate of the drug.

        2) Single and multiple dose pharmacokinetic trials - To assess the degree of linearity and drug accumulation.

3) Dose proportionality studies - To establish drug product performance over the range of doses and dosage forms.

4) Fed/Fasting studies - To assess the impact of a standardized high fat meal on the bioavailability of the final dosage form. (Usually done with the highest proposed to-be-marketed strength)

5) Linkage studies - To "link" the clinically studied formulations to those proposed for marketing. Linkage is accomplished by demonstrating bioequivalence between the formulations according to standard bioequivalency criteria.

6) Sub-Population Studies - To assess the impact of the following specific demographic factors or clinical settings on the observed pharmacokinetics:

   a) Age

      (1) An evaluation should be made to detect any PK differences in elderly (>65 yrs old) or very elderly (>75 yrs old) subjects.

      (2) Pediatric PK data should be collected if the agent is intended for use in patients <18 yrs old.

   b) Gender

   c) Hepatic Impairment - ***E***valuation in this subject group should include subjects with mild to moderate hepatic impairment as measured by any of the accepted staging methods of hepatic function (i.e., Pugh score, Child-Turcotte, etc.)

   d) Renal Impairment - The evaluation is concerned with potential PK alteration in

subjects with mild to severe renal impairment. (Because of the confounding effects of surgery upon renal clearance, the use of the renal transplant model for such studies is not recommended.)

e) Disease - In cases where PK data is first collected in healthy volunteers, adequate data should also be collected in patients with active disease to determine if the disease state causes alteration.

b. *in vitro* pharmacokinetic studies for all drugs

1) *In Vitro* Dissolution (for solid oral dosage forms)

a) Designed to assess product performance and manufacturing quality control.

b) Can in some instances be used as a predictive tool to assess rate of drug release via development of an *in vivo* X *in vitro* correlation.

2) P-450 Isoenzyme Studies - Can be used in a predictive way to evaluate drug-drug interactions that occur via induction or inhibition of specific metabolizing enzymes.

c. *in vivo/in vitro* pharmacokinetic studies for anti-rheumatic drugs.

1) Trials in women - As women are usually the larger target population for anti-rheumatic drug therapy, their representation in pharmacokinetic trials should reflect this.

2) Protein Binding - Because poly pharmacy is common during the treatment of rheumatic disorders, *in vitro* binding studies with blood from patients with active disease should be used as a preliminary screening tool for potential displacement reactions.

2. PK/PD DRIVEN CLINICAL TRIALS

Basic PK information is needed for product approval. Beyond this, the use of carefully collected blood levels and proper PK definitional studies can be a major guide to understanding and optimizing both dose-related efficacy and dose-related toxicity responses. Furthermore, there is a rarely tapped potential for the use of PK or PD “driven” trial designs to increase the efficiency of development. However, PK driven strategies (pharmacokinetically controlled designs) would be less efficient in circumstances where serum levels and clinical effects are only weakly linked. The most efficient scenario is the availability of a convincing PD bioassay (closely predicting clinical response and toxicity response) to use in PD controlled designs.

The discovery and application of a plasma marker of disease progression or severity would facilitate PD driven designs. None exists to date. In practice it is much more likely that inhibitory models of drug response, such as those used in organ rejection, will be the first to be developed.

C. Considerations in Phase 1 Trials

For general information on clinical development pertaining to most drugs and biological products, see "General Considerations for the Clinical Evaluation of Drugs."[5]

"Phase 1" has two distinct meanings: one meaning refers to the earliest, first-time-into-humans trial; while the other encompasses studies of pharmacokinetics, metabolism, drug interactions, special populations and other clinical pharmacology trials described above. It is expected that both kinds of Phase 1 trials will ordinarily be conducted during the clinical evaluation of therapies for RA. This section is primarily intended to discuss issues related to the initial Phase 1 trials, i.e., the first time people are exposed to the drug, or to a particular dose level, or duration of therapy.

1. Settings and investigators

First-time-into-humans Phase 1 studies should be carried out in institutions with a full range of clinical and laboratory facilities and the patients should be kept under close observation. It is desirable that the trials be under the direction of rheumatologists with experience in early drug development, or that a team of investigators combining experience in rheumatology and clinical pharmacology be employed.

2. Subjects

Traditionally, first-time-in-humans drug trials have been conducted in healthy volunteers. Such studies were predicated upon the ability to perform, and to interpret the results of, preclinical animal tests. If the preclinical testing did not reveal potential mutagenic, immune system or possibly serious or long term effects at or near the expected therapeutic range, testing in volunteers commenced. Since many biological products, and a number of categories of drugs (e.g., antineoplastics, immunosuppressives, many "DMARDS") have narrow margins of safety or therapeutic indices, initial testing has been done in patients. This has created challenges in selecting an appropriate initial patient population.

For drugs and biologics that have been tested in relevant preclinical toxicity evaluations and have been found relatively safe and are also not believed to have the potential for mutagenic, immune system or other serious or long term effects (e.g. based upon mechanism of action or product class specific effects) at doses at or near the therapeutic range, trials may be initiated in healthy volunteers. If however, significant effects have been demonstrated or might be possible (e.g., modification of immune function), selection of an appropriate patient population is necessary. It is recommended that patients have RA by ACR criteria. Patients should be without other serious medical conditions. Selection of the most appropriate severity level can be difficult. There is a need to balance risks, commensurate with anticipated benefits, against the need to avoid toxic agents in very ill patients. Therefore, the criteria used to select these patients should generally be more restrictive than those used in later studies to ensure that the patients have sufficiently severe disease to justify the risk of a relatively unknown agent and that they do not have other medical problems that increase the risks. Patients with

minimal disease are often not appropriate for the same reasons that the testing is not initiated in healthy volunteers. Patients with devastating RA also are not the best starting population: they frequently have medical complications of their disease and its treatment, and they may be less likely to respond to therapy. Patients who have failed to respond to a standard DMARD or who have had side effects requiring discontinuation should be considered candidates. In any case it is particularly important that informed consent be complete and that some provisions be made to assess that patients understand what they are consenting to. If the potential exists for disease exacerbation, this should be part of the informed consent.

When the characteristics of the agent suggest that it may potentially have long-term gonadal effects, it is desirable that men and women not wishing to parent children be chosen for Phase 1 studies.

3. Trial design

Ordinarily initial Phase 1 studies are sequential dose escalation trials, in which safety and tolerance at a specific dose is established before exposing additional subjects to a higher dose. A single dose is almost always tested first, followed by repeated dose studies; however, this design is influenced by the type of agent used. Although escalating the dosage to a clearly determined maximal-tolerated-dose (MTD) will aid future trial design, in some instances it is not medically prudent to try to fully characterize the MTD.

There are published suggestions (Harter, DMARD 1 proceedings) for dosing and escalation schedules for Phase 1 drug trials in RA. The starting drug dose chosen is typically a "no adverse effect" dose (determined by interspecies mg/kg/day dose conversion from animal to human). For biologicals, the initial dose chosen is often recommended to be one thought to have no adverse biologic effect. Conservative dose escalations (e.g., half log or less), have also been recommended.

4. Concomitant therapy

Use of low-dose corticosteroids (up to 10 mg prednisone equivalent daily), and NSAIDS may ordinarily be continued in Phase 1 trials. Concomitant therapy with methotrexate and similar agents should be avoided in initial phase I trials of all novel antirheumatic drugs, biologics and devices because of the difficulty of assessing toxicity of the novel agent and of the possibility of added or synergistic toxicity with coadministration. However, because physicians now prescribe methotrexate and similar agents earlier in the course of rheumatoid arthritis, it may be difficult to identify adequate numbers of patients not taking these agents for study. Approaches to allow the use of methotrexate and similar agents in later phase I trials include: (a) obtaining reassuring evidence of lack of toxicity from relevant animal models in which coadministration occurred; and (b) starting at doses much lower than the "no adverse effect level" as determined by preclinical studies. ***(Question for panel - Should a preclinical animal coadministration toxicology study be recommended?)***

5. Observations

   a. Safety. The standard batteries of safety observations are well described in many publications. However, depending on the characteristics of the agent, additional types of safety observations may be necessary, e.g., tests of effects on cellular and humoral immune function or host defenses. For products with the potential for effects lasting long after administration, or for delayed toxicity, appropriate follow-up should be designed. For example, Phase 1 studies of agents used to deplete or modify the function of T-cell subsets should be designed to carefully assess both the short and long-term effects on number and functional status (e.g., DTH responses) of cell populations and other pertinent pharmacodynamic assays during therapy and during follow-up.

   b. Efficacy. It is recognized that developing an understanding of the agent's therapeutic potential in early trials is highly desirable for efficient drug development. However, given the realities of evaluating RA responses in open trials, estimation of the agent's potential effectiveness in the earliest trials is fraught with pitfalls. Ordinarily, sponsors will

wish to evaluate "proof-of-concept" as part of Phase 1: i.e., does the agent have the pharmacological effect predicted from the preclinical development and intended to mediate the treatment effect?

D. Considerations in Phase 2 trials

During Phase 2, larger, often longer-term trials are employed to better define the dose- and exposure-related toxicity and activity of the agent and to explore its effectiveness. Enough information should be generated to ensure that the Phase 3 trials can be conducted safely and with a high probability of success. In addition, sponsors should solidify a total drug development strategy during Phase 2, to ensure that, after Phase 3 safety/efficacy trials are done, that all the information needed for registration will have been gathered, including an appropriate safety database, clinical pharmacology, dose response data, any needed exploration in special populations (e.g., renal failure, hepatic failure), drug interaction information with agents expected to be coadministered and so forth. Achieving this will frequently require the conduct of additional "Phase 1" clinical pharmacology studies and Phase 2 trials while the Phase 3 development is ongoing.

The following issues are important for Phase 2 trials in RA:

1. Dose finding.

This is a central challenge of Phase 2 development. Once a reasonably safe range of doses has been established, randomized, parallel arm dose-comparison trials are ordinarily recommended. If feasible, use of a placebo arm is desirable for several reasons. First, if no difference is found among doses, either all may be equally effective or equally ineffective. Second, if some trend to a dose-response is found, the placebo arm may give some "reality-testing" about the possible magnitude of the observed effect. If use of a placebo is not possible, designs that include an arm with a well-characterized, reasonably effective therapy as a active control can also be very useful.

For agents that are thought to have prompt action, and rapid ending of effect once stopped, alternative designs, including cross-overs[6] and even titration designs, may be useful, although these

have not been usable for traditional "DMARDS" because of their delayed onset and offset of action.

The desirability, even necessity, of identifying a range of doses with acceptable toxicity and reasonable activity, for study in Phase 3, cannot be stressed enough.

2. Safety

Many of the newer RA therapies being investigated raise somewhat novel safety concerns. Because of the potential for significant, long-lasting or delayed-onset toxicities, it is desirable to design the Phase 2 studies to provide a "leading edge" of exposure and longer-term follow-up, for the larger number of patients who will be exposed in Phase 3. Provision for long-term follow-up can be helpful in addressing concerns at registration, e.g. relating to the potential for immunosuppression, opportunistic infections, neoplasia, and induction of autoimmune disease.

3. Additional development aspects

a. Concomitant therapy. Before starting Phase 3 trials, some idea of the agent's interaction with other agents likely to be used by the target population should be developed. This could include be gleaned from knowledge of metabolic pathways, studies in in vitro systems, animal or human pharmacology studies, or drug interactions. The former types of information could help direct the sponsor to areas needing actual clinical evaluation. When products are intended to be tested as combination therapy with the investigational agent, substantial information on interactions and safety of co-administration should be developed in Phase 2.

b. Gender effects. Most RA trials have predominantly female enrollment. Sponsors are urged to evaluate whether the observed safety and efficacy findings are restricted to women or can be also extrapolated to male subjects, based on subset analyses from trials, PK data, or other information.[7] (Ref exec summary of gender workshop and "gender guideline")

E. Efficacy Trial Considerations

1. General considerations

FDA's experience in reviewing numerous RA trials has shown that there are recurring challenges in performing, analyzing, and interpreting trials of effectiveness. These problems are well known to RA trialists, but are worth summarizing here. The later sections on trial design, outcome measures, and trial analysis contain suggestions on how to deal with some of these problems.

a. The problem of dropouts. Many RA trials suffer from extensive dropout. Patients drop from the trial for the usual reasons of adverse effects, loss of interest, moving, etc; however, there are often substantial dropouts for lack of efficacy. Differential, or high levels of, dropout, either for adverse effects or for inefficacy, create tremendous analytical problems in the evaluation of efficacy, particularly for agents with modest treatment effects. Last observation brought forward (LOBF) intent-to-treat analysis is not a panacea for these problems (for example, patients who drop out for inefficacy may often continue to get worse, a trend not captured in the LOBF analysis.)

It is particularly important, therefore, that trials be designed with an eye toward preventing dropouts. Some tactics that may be used include:

1) Screening patients initially to eliminate those with little commitment to participate.

2) Using studies with run-in periods of various sorts, randomizing patients to treatment groups only after their eligibility is confirmed.

3) Choosing effective active controls in long-term studies.

4) Thoroughly training investigators so that enrolled patients are eligible, the protocol is not violated, etc.

5) Designing trials explicitly to maximize retention. Traditionally recommended RA trial designs have focused on eliminating sources of variability, for example, extra pain medications, intra-articular injections, etc. Often, patients whose treatment constituted a major protocal violation were dropped from the study. There is an explicit trade-off between patient retention and tolerance of variability that should be recognized in RA trial design. Protocols demanding rigid adherence may yield uninterpretable results because of patient and investigator intolerance of the requirements. On the other hand, protocols permitting any kind of additional intervention may likewise be so confounded as to defy interpretation. Sponsors should assess where the balance lies for the particular agent and use being contemplated.

6) Following up patients who have stopped treatment. Many patients who withdraw from the study can be evaluated for outcomes (for example, at the end of the study period) although they are no longer on active treatment. This assessment can help the evaluation of the impact of dropouts on the results.

Another way to deal with dropouts is to include them as part of the endpoint--for example, in a time to treatment failure analysis, or a by-patient success or failure endpoint.

b. Problems with control groups

Conduct of placebo-controlled trials in RA is often problematic, and becoming more so. Dropouts for inefficacy are frequent, and may lead to uninterpretable results. The growing trend toward earlier intervention in active disease means that investigators are less likely to be in "equipoise" as far as the appropriateness of the trial, especially if it is longer term.[8] In some cases the only acceptable group for placebo-controlled trials are patients who have failed multiple other therapies. These patients may have

particular disease characteristics that are not favorable for demonstrating a treatment effect.

Most placebo controlled trials in RA are conducted as "add-ons" to standard therapy, i.e., the new agent plus background therapy vs. background treatment alone. There is increasing interest in considering methotrexate one of these background treatments. This raises safety concerns, and also a trial design like this leaves unanswered the question of how the agent would perform added onto only NSAIDS and corticosteroids, the usual components of "background therapy".

A problem with active controlled trials in RA is that a number of the approved agents are of marginal efficacy. Comparative trials intended to show "equivalence" to such treatments, when not anchored by a placebo control group, often lack credibility, unless the intent is to statistically "beat" the approved treatment. The concern is that the marginally effective drug will fail to exert a detectable effect in the trial (surprisingly, a not uncommon outcome in clinical trials) and the new treatment will be declared "equivalent" to this non-effect. This problem of "assay sensitivity" appears to be nonintuitive and widely misunderstood.

In instances where it is possible to show a convincing dose-response on efficacy, such concerns are moot. However, this has proven to be very difficult with currently available agents.

c. Artifacts of trial designs

One of the more interesting findings in evaluating numerous RA trials has been the discovery of many artifacts resulting from specific trial designs. Some of these are not intuitively obvious. Consideration of the following may assist sponsors in evaluating their data and in trial design.

1) Withdrawal and flare designs. In an attempt to select patients with active disease, in some designs patients are withdrawn from their background

treatment (excepting corticosteroids) and allowed to "flare". Individuals with sufficiently high scores are enrolled into the trial. A number of observations have suggested that a very significant component of the observed "flare" is irrelevant to disease assessment. First, many patients randomized to the placebo arm of the trial rapidly return almost to their baseline state without further treatment. In fact, a return to near baseline state is a characteristic of the whole placebo population in these trials. Second, when patients undergo blinded withdrawal from therapy within trials, such dramatic flares are rarely observed, and certainly do not occur for the withdrawn population as a whole. It appears that there is expectation bias on the part of patients, who have been told about the flare procedure, and ascertainment bias on the part of investigators, who wish to have patients meet the entry criteria and enroll in the study. These introduce uncertainty and instability around the outcome measures used in such trials, and should be kept in mind when employing these designs.

2) Regression to the mean. A proportionately much smaller, but nevertheless noticeable and prompt "regression to the mean" is noted in the joint scores of patients required to have a certain minimum value for trial entry in trials not employing a "flare" strategy. This means that patients, on the whole, will not actually be as active as anticipated when the entry criteria are set. The mechanisms are similar to the above example.

3) Comparison to baseline in long term trials. One phenomenon frequently observed in RA trials is that patients who stay in the trials do better than those who drop out: "Responders do better than non-responders." This is true for both placebo groups and active treatment groups. In fact, after reviewing a number of these trials, an non-rheumatologist might conclude that the natural history of the disease is inexorable improvement. This fact makes

comparison-to-baseline outcome measures very difficult to assess, since all groups improve, and, as in the flare design, the bulk of the observed improvement may not be specifically treatment related.

2. Choice of population in efficacy trials

   a. It is recommended that patients enrolled in efficacy trials have RA as defined by ACR criteria, unless some other specific subgroup is targeted.

   b. Targeting subgroups of patients

      Taken together, many lines of evidence suggest that each rheumatic disorder, as defined by our current classification criteria, is actually composed of a number of more or less distinct diseases that cluster within a unit delineated by a common genetic background, corresponding clinical manifestations, and similar serologies, responses to therapy and prognoses. Therefore, the study and therapy of these diverse sign-symptom complexes should be enhanced by dividing them into more homogeneous groups defined by one or more of these common features.

      The increasing power of novel epidemiologic and molecular genetic approaches may lead to identification of even greater numbers of subgroups of most autoimmune and rheumatic diseases. Prospective studies are needed to confirm the clinical usefulness of some purported prognostic factors. Nonetheless, because therapeutic studies of the rheumatic diseases of necessity involve relatively small numbers of subjects, randomization alone may not ensure the comparability of treatment and control groups with respect to important predictors of prognosis. Thus, in those instances where data strongly support clinical, serologic or genetic markers as prognostic indicators, they should be taken into consideration in the design of trials, either through use as a covariate, through stratification, or by defining separate trials. In RA the presence of rheumatoid factor, erosive or vasculitic disease, and DR4 homozygosity

correlate with a poor prognosis, so it may be advantageous to design trials accordingly. Although in some cases such studies may limit generalizability and impact labeling of the final product, it is also possible that the improved risk/benefit considerations inherent in such targeting will allow for more optimal use of the product and greater evidence of efficacy.

Other characteristics of the population, such as degree of functional impairment or radiographic progression, should be determined based upon the desired claim, the characteristics of the product, and the degree of generalizability of results desired.

c. Activity of disease

In order to enhance the power of the trial, it is desirable to improve the chances of a response to therapy. Traditionally, enrolling patients who meet preset levels of disease activity, or those who flare on withdrawal of background treatment, have been used to enroll those with active disease. Although these methods can be successful, they also may introduce artifacts as discussed above. Measurements of outcomes in trials using these patient selection methods should be designed with these results in mind. Other maneuvers to improve responsiveness include using run-in periods to eliminate patients with poor compliance or highly variable disease.

Sponsors should consult with Agency personnel on the generalizability of claims derived from trials with significant limitations on entry criteria.

3. Concomitant therapies

Most patients in RA trials will be consuming a myriad of other medications. Use of medicines unlikely to influence treatment outcomes (e.g., antihypertensives) should simply be recorded, although investigators should be alert for possible drug interaction (e.g., with oral contraceptives). For handling the use of arthritis medicines or analgesics, the following approaches may be used:

a. Prohibition of their use. This strategy may result in noncompliance or an increased number of dropouts.

b. Incorporating protocol-specified use, with monitoring. With this strategy, additional analgesic use (and possible other arthritis medications) may be used according to protocol specified criteria.

c. Using the need to add analgesics, or their quantitative consumption, as an efficacy endpoint.

d. Using the need to add more arthritis treatments as an outcome measure, either one of a set or as an indication of "treatment failure."

As previously stated, it is desirable that there be some information providing a comfort level on drug interactions between the test medication and expected concomitant arthritis therapies.

4. Efficacy Trial Design

a. Choice of controls

The choice of controls in an RA trial will often be dictated by feasibility and ethics. Placebo-, dose-, concentration- or active-controlled designs are acceptable. It is desirable that at least one efficacy trial be designed to show an unequivocal treatment effect, i.e., the test drugs "beats" a randomized control arm, whether that be a lower dose of the agent, an active control, or a placebo.

b. Stratification

Although randomization is intended to balance, on average, known and unknown biases and confounders, in any specific trial, especially a "small" one, randomization may be grossly inefficient and fail to balance. "Small" here usually means less than hundreds of patients per arm so most foreseeable RA trials are "small" for these purposes. Rather than rely on randomization, it may be advisable to stratify with known (or highly suspected) major risk factors

(demographic, clinical--including prior treatment failure experience, serologic, genetic, radiographic) to ensure their balance across arms. A rule of thumb is that any factor as strong as the treatment group influence on the outcome should be considered a potential stratification factor.

However, there is a legitimate methodological debate as to whether extensive pre-trial stratification is preferable to leaving some factors for analytic post-stratification (i.e., as covariates). A highly stratified design will impact anticipated accrual in a major way, and many stratification factors in a trial showing only weak effects can be problematic because, through their joint partitioning, small or even empty cells may result in (for example, an ANOVA) making the results sensitive to small data perturbations. For these reasons it may be preferable to reduce stratification and use instead prospective agreement on factors as candidates for covariates in the analysis.

c. Blinding

Full patient and assessor blinding are necessary for a credible inference, unless the anticipated effect is much greater than anticipated bias/confounding. Partially unblinded designs are not optimal, but the strength of their inference will increase if the following conditions exist:

1) differential dropouts cannot occur,

2) unblinded treating or monitoring investigators are segregated from blinded assessing investigators, and

3) the primary endpoint is inaccessible to patients and investigators throughout the trial (e.g., an X-ray).

Designs often have compromised blinding unless either there is rough parallelism in time to onset, nature of response, and toxicity profile of the two agents being compared, or the design uses segregated assessors. Trials should have parallel dosing in both arms so that if a drug requires frequent dose manipulations the blind is not threatened.

d. Choice of outcome measures

Outcome measures should be selected to support the desired claim.

1) Measures to support claim of "treating signs and symptoms of RA": Assessment of joint counts: 66 or 28 joint count acceptable.[9] [10]

2) Remission induction: 1981 ACR Criteria: categorical score

3) Prevention of structural damage. Current perspective on studies in established RA:

Measurement of X-ray changes in RA has a substantial history, but many problems remain to challenge the trial designer - see a recent review (Scott: J Rheum 21(41):36-40,1994). The two major approaches used today are the modified Sharp (ref) and the Larsen (ref) methods. Methodologic work (Sharp A&R:28:16-24, 1985, Grindulis Rheum Int 3:39-42,1983) to date has only addressed reproducibility, wherein (if readers are properly coached) both methods perform well and show sensitivity, with one study suggesting the Sharp method is more sensitive (Cuchacovich A&R:35:736-9,1992). The discriminatory power for detecting change in X-rays was found (Sharp, 1985) to be approximately one percent of the total score (0-150 for the Larsen, 0-314 for the modified Sharp). It has been argued that "joint space narrowing," the other dominant X-ray feature assessed in addition to erosions, does not yield any additional information over that made from quantifying erosions alone, as evidenced by high correlations found between measurements of erosions and joint space narrowing.

Newer imaging techniques as possible RCT trial endpoints: The emergence of computer enhancement of plain films and of CT and MRI technology already is appearing to allow

much more sensitive quantitation of X-ray changes (such as erosions) and quantitative visualization of other joints structures (e.g. cartilage volume, synovial load), and future developments may even yield indicators of pathophysiology rather than simply anatomic measures. Many of these will have a convincing basic science claim as a valid endpoint for an interventional RCT in RA, but unless early studies are dramatic (e.g. clinical cures) the logic requiring a process of characterization and determination of their clinical relevance will exist for these markers too, invoking again some moderately long-term validating database.

4) Prevention of disability

(**Question: suggestions for appropriate measures?)**

5) Health status and QOL

(**Question: suggestions for appropriate measures?)**

5. Other design issues

a. Defining Criteria. It is of critical importance that the outcome measures for the study be clearly and precisely defined, and that the criteria for a study "win" also be precisely defined, including the statistical test that will be applied. Efficacy trial protocols should contain an analytical plan that lays out the primary comparison(s) to be made, the test for "success" of the trial, and the statistical techniques intended to be applied. The sponsor should also be clear about what claim success in the trial is intended to support.

b. Handling of missing data and dropouts. The proposals for handling missing data, including dropouts, should be clearly specified in the analytical plan. Methods that do not require imputation of data are preferable. Some of these are described in the following section on statistical techniques.

c. Comparative trials. In trials employing active controls, an effort should be made to ensure a "fair comparison", whether the objective of the trial is "beating" the comparator or showing equivalence. An appropriate dose of the active control should be chosen, as well as a patient population in which the comparator is generally indicated. In particular, patients who have previously "failed" the particular comparator are not ordinarily considered appropriate for enrollment.

d. Trials intended to demonstrate equivalence between the test agent and an active control. A central problem in trials of equivalence is ascertaining whether the active control actually had a measurable beneficial effect. It is desirable in equivalence designs that highly effective comparators be chosen, and that they be used in the optimum dose and patient population, as above. The criteria for determining equivalence should be prospectively stated. Standard confidence limit approaches may be used.

6. Statistical Considerations in Efficacy Trial Design

It is generally a good idea to discuss the design and analysis with the FDA review team prior to embarking on a study. We recommend that sponsors familiarize themselves with the Guidelines for Statistical Analyses (CDER, 1988.)

a. Efficacy Endpoints

The goal of the statistical analysis of the endpoint Is to demonstrate (or not) if the product gives convincing evidence of efficacy. Studies of RA are generally longitudinal (i.e., measurements are taken at several times). This implies that appropriate statistical methods be employed to analyze these data. Since the 1960's there has been much statistical research done in this area including Growth Curves. Repeated Measures, Random Coefficient Regression, and GEE models. Statistical routines are available in most major statistical analysis programs to perform these.

The endpoints should be specified in the protocol for the study and the proposed analysis should be outlined. In the analytical plan, the method of determining the sample size should be stipulated in sufficient detail to permit the FDA to verify the computation. It is useful to indicate how the assumed treatment effect is obtained. There are several options available to the sponsor. The response may be a binary variable which indicates improvement from baseline. Such an endpoint has a straightforward interpretation and analysis if all patients are included in the study to its completion . If some patients have only partial follow up, there is considerable question about how they should be scored, failure, status at last follow up, least favorable score for the approval of the product. Whatever procedure is adopted, it should be specified in the protocol and the analysis discussed. A second class of responses might be ordered categorical responses (e.g., much worse, worse, no change, better, much better). Such responses are usually analyzed using ranks (accounting for ties). This leads to a Wilcoxon rank-sum test. The response is measured at the specified ending time of the patient regimen. Again, if patients fail to complete the regimen, an appropriate final score should be assigned. If the endpoint is a continuous response (e.g., time to an endpoint, tender joint count), the difference between final and baseline may be used ("change score"). This is a widely used method - it has the advantage of measuring a difference of general interest. It does not account for time. If one divides the change score by the time interval, a rate of change per unit time arises. This allows inclusion of all patients whether they complete the study or not. A similar method would fit the slope and intercept for each patient's measurements over time (i.e., time is the predictor variable). The slope and change rate are highly correlated for small numbers of observations.

In general, the appropriateness of the statistical model should be assessed in the analysis of the studies. These diagnostic measures might include checking for outliers determining if distributional assumptions (usually normality) are met if common variance assumptions hold (homoscedascity), and if the model is correct. In RA, a

problem might be due to a delay in the effect of the drug. Thus, a straight line model of improvement might be displaced from the start of therapy by several months. In this case an AUC or mean-response analysis might be the simplest.

Another aspect of RA studies is that multiple measures are used to ascertain disease status. A common method is to require the product to "win" on three of four measures. An alternative would be to use multivariate analysis methods. Methods for one-sided alternatives have been studied (as far back as 1969, but also more recently by Follman). It is easy (although cumbersome) to construct the likelihood ratio test for this hypothesis. As with all such methods, there is concern over the effect of non-normality on the test. Sponsors may select a multivariate endpoint (e.g., tender joint count, swollen joint count, patient global assessment) and analyze each separately. In such cases, the multiple testing aspects of these endpoints should be adjusted for. One simple method is to use a Bonferroni adjustment, another would be to perform a global test using a multivariate procedure (such as the T2 test) followed by tests of individual variables only if the overall test is significant.

As noted above, patients who fail to complete the prescribed follow up create serious concern at the time of analysis. Many methods have been proposed. Last Observation Carried Forward ("LOCF") simply gives the subject the score they received at their final visit.

This maintains the change score, but changes any rate of change computations. Multiple imputation replaces the missing values with 'reasonable' ones which have a distribution similar to the known sample. If this is done several times, an estimate can be obtained of the variability induced by the replacement procedure. Fitting a line to the response and obtaining the predicted value at the final follow up time allows the sponsor to compare results at the final time. Finally, survival methods might be used. These would analyze the time to dropout or achieving a specified

level of the response variable - e.g., time until dropout or tender joint count increased by 10. All of these suffer from the problem of informative censoring. That is, the patient's failure to complete the study is likely related to their response score. One partial adjustment for this might be to use an indicator variable (no dropout or dropout) as a covariate.

b. Composite Endpoints - items such as the ACR criterion for improvement are proposed as acceptable endpoints. These create special analytic problems since they are a series of O-1 variables comparing to baseline. If only the final value is used, the interpretation is clear. If there are several assessment times, a growth curve of binomial variables is required. This can be handled using GEE models.

c. Individual Endpoints - measuring improvement on the individual components of the ACR criterion may be handled with the methods mentioned in the Composite Endpoints paragraph. Alternatively, these might be analyzed as scored (e.g., tender joint count difference or slope). In this case, rank methods are frequently useful as they are more robust than the normality based statistics.

d. Remission Endpoints - The time to a durable remission (which should be defined by the sponsor and accepted by the FDA reviewing group) may be analyzed using survival analysis techniques. The proportion achieving a given decrease in a joint count could be analyzed using logistic regression models.

For some endpoints, there is a non-linear response; e.g., the natural history of the disease is for the endpoint to increase for a time, and then decline. In this situation, if a protocol specified patients early in their disease, all products might appear to have little or no effect. If late stage patients, all therapies might look terrific. In such cases, it is prudent to develop and validate a model for the response in the phase I and II studies and then examine the effects on the parameters of the model in the phase III studies.

e. Site effects - If the patients have been stratified and randomized by site, the analysis should include a site effect. This removes this source of variation in the data. The analytic principle is that one should account for the experimental design in the analysis. There may also be a site by treatment interaction indicating that the treatment does not work equally at all sites. This may be the case when the sites have great variation in the number of patients they treat.

F. Safety Analysis

The approach to evaluating adverse reaction data and laboratory values has traditionally differed from that used to evaluate efficacy. The purpose of safety evaluations is rarely to test a specific hypothesis, but rather to examine the pattern of effects and to detect unusual or delayed events. Analyses using cumulative occurrences, scatter-plots of laboratory values (baseline vs on-therapy), or general regression techniques may be helpful. The safety profile should address to what extent adverse events (drug reactions or lab values) depend on duration of drug exposure, dose level, coexisting medical conditions, or possible drug interactions. Incidence rates should be calculated using denominators that reflect the period of drug exposure for the population at risk. Cumulative incidences (hazard rates, instant probabilities) better represent the temporal pattern of drug effects than do prevalence rates, and comparative cumulative incidence tables - drug vs active control(s) vs placebo - are very helpful to clinicians interpreting the label.

An attempt should be made to characterize the patient population susceptible to adverse drug effects. Some extraneous factors can complicate the safety data, such as variations in soliciting and reporting adverse reactions among the investigators, and differences in the definition of normal ranges for lab values among different laboratories. Since adjustment for their effects may be difficult, precautions should be taken in the design stage of the trial to minimize the influence of these factors by preparing clear and specific instructions for data collection, and monitoring adherence of the investigators and the laboratories to the protocol.

III. SPECIAL CONSIDERATION FOR BIOLOGICAL PRODUCTS

Although there are similarities between RA trial designs for drugs and biologics, biologics have certain characteristics and present certain problems that should be considered in their development.

A. Species Specificity

The schemes used traditionally in determining the initial human dose may not pertain to biologics. Biologic agents may behave differently in animal models than in humans, depending on the physiologic relevance and avidity for the receptor of the ligand in the animal compared to the human.

B. Dose Responses

The dose response curve may be steep (narrow therapeutic window) and/or even hyperbolic, and an agent can be quite toxic at levels just above those thought to show efficacy.

C. Toxicity Response

The toxicity response curve may be highly unpredictable and potentially very dangerous, and include the risk of disease worsening. Biologics may have the potential for disruption of immunologic and physiologic processes. Monoclonal antibodies to cellular epitopes of the immune system, for example, or to TNF receptors can or may cause serious morbidity at doses only slightly higher than those that are efficacious with markedly less toxicity.

D. Product Homogeneity

This often plays a critical role in activity and toxicity of a compound. Product alterations can greatly affect physiologic activity. Thus, biologics should have consistent lot-release criteria and be reasonably well characterized to be properly evaluated.

IV. SPECIAL CONSIDERATIONS FOR MEDICAL DEVICES

A. Background

Medical devices for the treatment of RA are diverse in therapeutic effect, ranging from agents designed for primary therapeutic effectiveness to

those utilized as therapies adjunctive to drugs or biological agents. The diversity is also due, in part, to the heterogeneity of RA disease activity and variability in its clinical manifestations. Although this problem is present for drugs and biologics, it is more obvious with medical devices. Device pre-clinical testing requirements cannot be generalized because devices for RA have a diverse range of chemical, mechanical, and electrical properties. In addition, the issues of the optimal placebo control and of local versus systemic effects are very common in the evaluation of medical devices. These factors are relevant to both efficacy and safety determinations as described below.

B. Efficacy Considerations

1. Some medical devices intended for local administration may have unexpected systemic therapeutic effects. Therefore, precise determinations of mechanisms of action should be made to minimize this phenomenum.

2. Use of a "sham" device is the most desirable placebo control for medical devices, but the success of patient and/or physician blinding with sham devices is not always be adequate. Patient and/or physician blinding may not be feasible if the product is delivered in a surgical or invasive medical procedure. Inadequate blinding usually biases efficacy determinations in favor of therapy. ***Therefore, monitoring and characterization of blinding methodologies is imperative.***

3. For devices intended to be utilized as adjunctive therapies to drugs or biologics, disease status and severity should be consistent and well characterized to minimize biases in endpoint outcomes. Similarly, the primary therapy with drug or biological agent should be consistent to avoid outcome bias, as should all additional adjunctive interventions such as hot/cold therapy, splinting, passive/active resistive therapies, ROM, energy conservation techniques, special orthotics, etc.

4. The issue of quality of life (QOL) determinations is very important for devices intended more for rehabilitative utilization rather than primary therapy. These determinations are also important because of the technical demands of certain device uses. QOL benefits should be judged in the context of ease of administration and

relative convenience of administration of devices, so provisions should be made to assess the satisfaction with therapy and the improvement in QOL. The outcomes of these determinations should be blinded from the participating investigators to avoid outcome assessment bias.

5. For devices necessitating in-hospital or in-office use, it is recommended that clinical utility be determined accurately and early in development. In addition to adverse event risks, the practical "risks" of the product, such as inconvenience or pain with administration, should also be characterized and judged as efficacy outcomes. Although it is very difficult to gather reliable efficacy data, let alone clinical utility, early on, this is critical for the sponsor in order to be able to make a reasoned “go/no go” decision. Agency consultation is advisable.

C. Safety Considerations

1. The availability of well-characterized short-term adverse event rates (cumulative 3-month incidence of about 1%), as described for drugs, may not be feasible for medical devices. Due to the more technically demanding administration of devices, it is generally not feasible to enroll large numbers of patients or to conduct several concurrent studies. The timing of device adverse events may differ from that of drugs in that common adverse events may not occur most frequently within the first few months of treatment. Therefore, patients with devices which have a delayed effect noted in preclinical or phase II testing should have extended follow-up beyond time on device. These factors may constrain the ability to capture adverse events needed to build an adequate safety database, and may therefore need to be addressed in post-approval studies designed to increase the duration of follow-up or increase the numbers of patient exposures.

2. Because some medical devices are administered in conjunction with a medical or surgical procedure, the distinction between a device-related or procedure-related adverse event is sometimes obscure. The nature, timing, and degree of severity are some factors used to help determine whether an adverse event is device- or procedure-related. These determinations are often based on clinical judgement, so if blinding is inadequate a potential for bias

exists. For this reason, the evaluator should be blinded to treatment (i.e., segregated treating and evaluating physicians). It is recommended that sponsors delineate in the protocol guidelines for assessing procedure-related versus device-related adverse events.

3. Although some medical devices (e.g., those emitting radiation or those administered with a procedure) for RA treatment may be used intermittently, some may be intended for chronic use, so identification of a maximum lifetime exposure or a maximum frequency of exposure to the device is important.

V. SPECIAL CONSIDERATIONS FOR JUVENILE RHEUMATOID ARTHRITIS

A. Background

Juvenile rheumatoid arthritis (JRA) is a heterogeneous group of diseases which share the common feature of chronic, idiopathic inflammatory synovitis, with onset prior to 16 years of age. These disorders have been divided into clinically distinct subsets based on the extent of joint involvement and extra-articular manifestations: pauci-, poly-, and systemic-onset JRA, as well as oligoarthritis associated with HLA-B27, and they have been further subdivided based on clinical courses[11]. Immunogenetic subsets appear to correlate with these clinical course subsets[12]. Of these various entities, only one subset, rheumatoid-factor positive polyarticular JRA, appears to be parallel to adult rheumatoid arthritis. For this subset, only pharmacokinetic studies are necessary to determine pediatric dosing (21 CFR pt 201). For all other JRA subsets, distinct clinical trials, separate from adult RA patients should be conducted. The heterogeneity in JRA, and the state of development of JRA assessment instruments, both have had a widespread impact on JRA drug development as noted below.

A heightened sense of strict ethical guidelines is brought to the design and conduct of trials in children, because pediatric subjects constitute a vulnerable population. These are published guidelines addressing the ethical conduct of studies to evaluate drugs in pediatric populations,[13] [14] [15] and all should be consulted. Clinical trials for patients with JRA should involve certified pediatric rheumatologists or adult rheumatologists with extensive training in pediatric rheumatology and demonstrated competence in caring for children with rheumatic disease. FDA

consultation should be sought early in the design process. Many investigational therapeutic agents for JRA will be treated as orphan drugs, because of the small size of these populations.

As a general principle children should not be subjected to an agent that has not been first tested for safety in adults. Testing may begin in children if the anticipated benefits based on existing knowledge may justify the anticipated risks. If the agent has potential for use in both adult RA and JRA, then PK-PD and initial Phase I data (including maximum tolerated dose) should be available for adults prior to testing in children. Testing of JRA drugs expected to be similar to existing drugs, and which do not represent a major advance or alternative to the basic mechanism, can be delayed until there is extensive efficacy and safety data from adults or other pediatric populations. An agent developed specifically for JRA (e.g., a biologic agent against a specific etiopathogenic focus unique to JRA, not present in adult RA) may need to be tested initially in children, because exposure of adult RA patients or healthy adult volunteers may be unrevealing.

B. Outcome Variables/Claims

Efficacy in JRA is similarly divisible into reduction in the signs and symptoms of RA, remission, prevention of structural damage, prevention of disability, and improvement in quality of life, as with adult RA. Measures of disease activity for JRA are currently under study and development. There is experience with certain variables, some representing modifications from adult RA, and some unique to JRA. Since JRA can be painless[16], tender joint counts and assessment of pain cannot be utilized as in adult RA[17]. The assessment of pain should be sensitive to the cognitive perceptions of pain by children of varying ages; use of an instrument which has been validated in JRA patients, such as the Pediatric Pain Questionnaire[18] is necessary. Assessment of physical function and quality of life should be done using standardized instruments which have been validated for use in JRA, such as the Childhood Health Assessment Questionnaire[19] or the Juvenile Arthritis Functional Assessment Report.[20]

Measures of disease activity and drug efficacy should address the characteristics of each subset, and include extra-articular manifestations (e.g., anemia, thrombocytosis, fever, rashes, polyserositis, organomegaly, or uveitis) and effects on local and systemic growth. Responder indices

for JRA are under development and should be incorporated in RCTs, and these composite indices will likely be needed to optimally assess JRA patients. Each will be subset specific; for example, many patients, especially those with pauciarticular JRA, will have a normal ESR, no joint pain or tenderness, and no fatigue when the disease is active.

Definition of long-term claims of remission or prevention of disability are yet to be developed for JRA. As a result, the criteria used for adult RA may be modified for JRA, with individualization for each JRA subset. For patients with systemic JRA and some polyarticular disease, resolution of all extraarticular disease is also necessary to declare remission. Because approximately 20 percent of JRA patients eventually develop spontaneous remission, remission claims for JRA should always be the result of placebo controlled trials of at least 12 months duration in which remission is sustained over six months and is at least 20 percent higher in absolute value than the observed rate in controlled patients.

Regarding claims for prevention of structural damage, although the Larsen and Sharp indices have not been validated in JRA patients, radiographic evaluation should include joint space narrowing and localized growth disturbances.[21] [22] [23] Prevention of new erosions and use of other sensitive measures such as MRI is applicable from I. C. of this document (pgs 3 and 4). Radiographic claims should be structured similar to adult RA based on comparison of films taken greater than or equal to one year with retention rate of at least 85 percent.

C. Sequence of Studies in Children

Phase I. Earliest studies should be single dose, dose escalation studies intended to obtain pharmacokinetic-pharmacodynamic (PK-PD) data and information regarding the biologic effect, followed by multiple dose, dose ranging studies with step-up to, or approaching, the adult recommended dose, if adult data are available. If adult PK-PD data are available, and if the coefficients do not differ significantly for adults and children, then the number of time points for specimen collection can be minimized yet still confirm the PK curves observed in adults. Whether separate PK-PD studies are necessary for each JRA subset is an individualized decision, but it may be advisable for systemic-onset disease because of the known greater drug toxicities seen with this subset of JRA.[24] [25] [26]

Phase II/III. Recommendations for efficacy studies are based upon the nature of the agent under development. Sufficient numbers of patients should be enrolled to ensure adequate evaluation of efficacy and safety of the product. Separate trials for each JRA subset are recommended, or, alternatively, a single, large trial with stratified enrollment may allow reliable conclusions about efficacy and safety for each subset. “Escape clauses" are necessary to permit children not responding to experimental therapy to revert to conventional or alternative treatment.

In trials of agents for short-term claims only (not long-term or remission claims), the decision of what control to use is a judgment based on expected benefits vs. anticipated risks and the current knowledge of the product. If only an active control group is used, historical "assay sensitivity" should exist, and the definition of, and statistical test for, establishing similarity should be specified. If no prior adult studies exist, or if the agent has a novel mechanism of action or represents a new class of drug, then a randomized trial is indicated. Open label extensions to obtain additional data about risk and persistence of benefit are very valuable. In trials of agents seeking a long-term claim for retarding disability or inducing remission, a placebo-control design (all patients receiving “background therapy”) is needed because of fluctuations in disease activity and a relatively high proportion of spontaneously remitting patients, especially with pauciarticular and systemic-onset disease.

Choice of patients: JRA patients are not normally deemed eligible for entry into trials of biologic agents, or other immunoactive agents, unless they have failed to respond adequately to at least one standard slow acting agent (including methotrexate at a dose of at least 10 mg/M2/wk), but there may be legitimate arguments for exception to this. If there is sufficient evidence that greater efficacy or less toxicity can be obtained using an agent in early or mild patients, then the need for previous slow acting agent failure may not be present. If there is a strong rationale for targeting a biologic to a specific patient subset, the trials should (at least initially) be conducted only in that subset of JRA with postulated efficacy. Ethical principles do not permit exposing other children to a potentially toxic agent if there is no anticipated benefit.

Longer-term follow-up and Phase IV studies. Whether or not patients continue to receive the agent upon discontinuation from protocol, attempts should be made to monitor them for an extended period,

including effects on skeletal growth, development, behavior, sexual maturation, reproductive capacity, and secondary malignancy.

D. Concurrent Antirheumatic Agent Administration

The goal is always to minimize concurrent therapies so as to limit its potential to confound the efficacy and safety results. However, limitations of concurrent medication cannot prohibit ethically justified treatments nor should it make the protocol so unattractive to parents, physicians, and patients that enrollment is threatened. If background treatment is necessary, early tolerance studies, to ensure safety of coadministration, will need to precede any large trials.

Concurrent medications are usually important prognostically and may merit stratification. In general, the principles regarding concomitant therapy outlined in Part V-C are also applicable. The doses of concurrent slow acting drugs or of prednisone should have been stable prior to study entry, and preferably remain stable throughout the duration of the trial. Intra-articular steroid injections ideally should be discontinued for a minimum of one month prior to beginning experimental therapy, or if that is not possible with polyarticular disease, assessment adjustments will be necessary, including, at a minimum, discounting the particular injected joint.

E. Multicentered Trials

Since the prevalence of JRA is low, trials that require large numbers of patients will likely have to be multicentered. Multicentered studies should employ a standardized protocol and data collection forms among all centers. Pretrial meetings of all investigators and other involved personnel are strongly encouraged to assure uniformity in protocol interpretation, patient evaluation, and data recording. Comments in the Analysis section on multicentered trials and center effects are applicable here.

REFERENCES

1. Felson, DT, et al, "The American College of Rheumatology Preliminary Core Set of Disease Activity Measures for Rheumatoid Arthritis Clinical Trials. 1993. Arthritis & Rheumatism Vol. 36, No.6: 729-740.

2. Felson, DT, et al, "American College of Rheumatology Preliminary Definition of Improvement in Rheumatoid Arthritis." 1995. Arthritis & Rheumatism Vol. 38, tp48: 1-9

3. Pinals, RS, et al, "Preliminary Criteria for Clinical Remission in Rheumatoid Arthritis." 1981. Arthritis and Rheumatism, Vol. 24, No.10:1308-1315.

4. Mukherjee, A., Chen, C., Jean, L., and Coutinho, C.B. (1993). "Preclinical pharmacodynamics of anti-inflammatory drugs." In: Integration of pharmacokinetics, pharmacodynamics and toxicokinetics in rational drug development, edited by Yacobi et al., Plenum Press, New York, pp 105-113.

5. Center for Drug Evaluation and Research, General Considerations for the Clinical Evaluation of Drugs (FDA, Washington, D.C.)

6. Furst, DE, Caldwell, JR, et al, "Serum concentration and dose-response relationships for carprofen in rheumatoid arthritis." 1988. Clin Pharmacol Ther. Vol 34 N.2, 186-194

7. "Guideline for the Study and Evaluation of Gender Differences in the Clinical Evaluation of Drugs," FR Vol 58, No.139

8. Freedman: NEJM 317:141-145, 1987

9. Smolen, JS, et al, Validity and Reliability of the Twenty-Eight-Joint Count for the Assessment of Rheumatoid Arthritis Activity. 1995. Arthritis 7 Rheumatism Vol 38, No. 1: 38-43.

10. Prevoo, MLL, et al, "Modified Disease Activity Scores that Include Twenty-eight-joint Counts, Development and Validation in a Prospective Longitudinal Study of Patients with Rheumatoid Arthritis." 1995. Arthritis & Rheumatism Vol 38, No. 1: 44-48

11. Cassidy JT, Levinson JE, Bass JC, et al. 1986. A study of classification criteria for a diagnosis of juvenile rheumatoid arthritis. Arthritis Rheumatism 29: 274-281.

12. Nepom B. 1991. The immunogenetics of juvenile rheumatoid arthritis. Rheumatic Disease Clinics of North America 17: 825-842.

13. Guidelines for the ethical conduct of studies to evaluate drugs in pediatric populations. Committee on Drugs, American Academy of Pediatrics. 1995. Pediatrics 95: 286-294.

14. Committee on Drugs American Academy of Pediatrics. General considerations for the clinical evaluation of drugs in infants and children. Report to the FDA by the Committee on Drugs, American Academy of Pediatrics. Available through the Freedom of Information Office of the FDA. FDA Contract No. 223-79-3003 & 223-82-3009, 1982.

15. Informed consent, parental permission, and assent in pediatric practice. Committee on Bioethics, American Academy of Pediatrics. 1995. Pediatrics 95: 314-317.

16. Sherry DD, Bohnsack J, Salmonson K, Wallace CA, and Mellins E. Painless juvenile rheumatoid arthritis. 1990. J Pediatrics. 116: 921-923.

17. Giannini EH, Lovell DJ, Felson DT, and Goldsmith CH. Preliminary core set of outcome variables used in JRA clinical trials. 1994. Arthritis and Rheumatism. 37 (suppl): S428.

18. Varni JW, Thompson KL, Hanson V. The Varni/Thompson Pediatric Pain Questionnaire. I. Chronic musculoskeletal pain in juvenile rheumatoid arthritis. 1987. Pain. 28: 27-38.

19. Singh G, Athreya B, Fries JF, Goldsmith DP. Measurement of health status in children with juvenile rheumatoid arthritis. Arthritis Rheum. 1994; 34: 873-880.

20. Howe S, Levinson J, Shear E, et al. Development of a disability measurement tool for juvenile rheumatoid arthritis. The juvenile arthritis functional assessment report for children and their parents. Arthritis Rheum. 1991; 34: 873-880.

21. Cassidy JT and Petty RE. Textbook of Pediatric Rheumatology. 1995. 3rd ed. WB Saunders. Philadelphia.

22. Lang BA, Schneider R, Reilly BJ, Silverman ED, and Laxer RM. Radiologic features of systemic onset juvenile rheumatoid arthritis. 1995. J Rheumatology. 22: 168-173.

23. Wallace CA and Levinson JE. Juvenile rheumatoid arthritis: Outcome and treatment for the 1990s. 1991. Rheumatology Clinics of North America. 17: 891-905.

24. Silverman ED, Miller JJ, Bernstein BB and Shafai T. Consumption coagulopathy associated with systemic juvenile rheumatoid arthritis. 1982. J Pediatrics. 103: 872-876.

25. Hadchouel M, Prieur A.M., and Griscelli C. Acute hemorrhagic, hepatic, and neurologic manifestations in juvenile rheumatoid arthritis: Possible relationship to drugs or infection. 1985. J Pediatrics. 106: 561-566.

26. Kornreich H, Malouf NN, and Hanson V. Acute hepatic dysfunction in juvenile rheumatoid arthritis. 1971. J Pediatrics. 79: 27-35.